AMENDMENTS TO THE CLAIM

1-21. (Cancelled)

22. (New) A method executed by an interface unit for putting a client on hold, the method comprising:

(a) intercepting, by an interface unit, a request from a client to access a requested server;

(b) determining, by the interface unit, that a current response time of the requested server exceeds a threshold;

(c) identifying, by the interface unit, the client as on-hold in response to the determination;

(d) establishing, by the interface unit, a waiting time for the client; and

(e) transmitting, by the interface unit, an on-hold request to an on-hold server based upon the waiting time.

23. (New) The method of claim **22**, wherein the response time is estimated, by the interface unit, from a recurrence relation

$$t'_{(i+1)} = \frac{(i-1)t_{(i-1)} + it_i}{2i-1} + (t'_i - t_i)K$$

where t_i denotes the response time at the i^{th} episode, t'_i denotes the estimated response time at the i^{th} episode, and K is a constant of error correction learned from ongoing traffic.

24. (New) The method of claim **22**, wherein step (b) comprises evaluating, by the interface unit, if the determined response time exceeds a guaranteed client-server response time established by the requested server.

25. (New) The method of claim **22**, wherein step (d) comprises determining, by the interface unit, an approximate waiting time for the client based upon the estimated current response time of the requested server.

26. (New) The method of claim **22**, wherein step (d) comprises delegating, by the interface unit, establishment of the waiting time to a code on an on-hold page provided to the client, the code corrects the waiting time based upon a round trip time and a response time provided by the interface unit.

27. (New) The method of claim **22**, wherein step (d) comprises providing, by the interface unit, a code to the client, the code receives a preferred wait time or on-hold preference from a user of the client.

28. (New) The method of claim **22**, wherein step (c) comprises selecting, by the interface unit, the on-hold server from a plurality of on-hold servers based upon the waiting time or an on-hold preference.

29. (New) The method of claim **22**, wherein step (c) comprises generating, by the interface unit, an on-hold request for a web page of the on-hold server.

30. (New) The method of claim **22**, wherein step (c) comprises identifying a web page from a plurality of web pages, each of the plurality of web pages providing different content according to different wait times.

31. (New) The method of claim **22** further comprising maintaining, by the interface unit, the client on hold until the response time of the requested server is less than a desired response time specified by a user of the client.

32. (New) The method of claim **22** further comprising:

receiving, by the interface unit, an indication that the user of the client is finished with the on-hold server; and

taking the client off on-hold.

33. (New) A system for putting a client on hold, the system comprising:

an interface unit intercepting a request from a client to access a requested server,

determining that a current response time of the/a requested server exceeds a threshold, and

identifying the client as on-hold in response to the determination, wherein

the interface unit establishes a waiting time for the client and transmits an on-hold request to an on-hold server based upon the waiting time.

34. (New) The system of claim **33**, wherein the interface unit estimates the response time from the recurrence relation

$$t'_{(i+1)} = \frac{(i-1)t_{(i-1)} + it_i}{2i-1} + (t'_i - t_i)K$$

where t_i denotes the response time at the i^{th} episode, t'_i denotes the estimated response time at the i^{th} episode, and K is a constant of error correction learned from ongoing traffic.

35. (New) The system of claim **33**, wherein the interface unit determines an approximate waiting time for the client based upon the estimated current response time of the requested server.

36. (New) The system of claim **33**, wherein the interface unit delegates establishment of the waiting time to a code on an on-hold page provided to the client, the code corrects the waiting time based upon a round trip time and a response time provided by the interface unit.

37. (New) The system of claim **33**, wherein the interface unit provides a code to the client, the code receives a preferred wait time or on-hold preference from a user of the client.

38. (New) The system of claim **33**, wherein the interface unit selects the on-hold server from a plurality of on-hold servers based upon the waiting time or an on-hold preference.

39. (New) The system of claim **33**, wherein the interface unit generates an on-hold request for a web page of the on-hold server.

40. (New) The system of claim **33**, wherein the interface unit identifies a web page from a plurality of web pages, each of the plurality of web pages providing different content according to different wait times.

41. (New) The system of claim **33** wherein the interface unit further maintains the client on hold until the response time of the requested server is less than a desired response time specified by a user of the client.

42. (New) The system of claim **33** wherein the interface unit further receives an indication that the user of the client is finished with the on-hold server, and takes the client off on-hold.